International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Deep Learning for Lung Cancer Classification from Computed Tomography Images: A Review of Architectures, Datasets, and Validation Challenges

Ramya S

Research Scholar, Department of Computer Science, Mangalore University, Mangaluru, Karnataka

***Abstract*—Deep learning has been applied to computed tomography (CT) image analysis for lung cancer classification, pulmonary nodule analysis, and segmentation. This review summarizes representative approaches based on convolutional neural networks, transfer learning, attention mechanisms, Transformer models, hybrid CNN–Transformer architectures, and segmentation networks. The review focuses on datasets, task definitions, reported metrics, validation procedures, explainability, and limitations affecting reproducibility. Lung-EffNet, an EfficientNet-based model evaluated on the IQ-OTH/NCCD dataset, reported 99.10% test accuracy and a receiver operating characteristic value between 0.97 and 0.99. Wavelet U-Net++ reported a mean Dice coefficient of 0.936 and mean intersection over union of 0.878 for lung-nodule segmentation on LIDC-IDRI. These results are study-specific and should not be treated as directly comparable because the datasets and experimental protocols differ. A 2025 systematic review and meta-analysis included 315 studies; 209 studies addressing lung-cancer diagnosis reported pooled sensitivity of 0.86, specificity of 0.86, and area under the curve of 0.92. The reviewed evidence indicates the need for patient-level data splitting, external validation, transparent reporting, uncertainty estimation, calibration, and clinically meaningful evaluation. High performance on an individual public dataset should not be interpreted as proof of clinical readiness.**

***Keywords*—Computed Tomography, Convolutional Neural Network, Explainable Artificial Intelligence, Lung Cancer Classification, Lung Nodule, Medical Image Analysis, Transformer.**

I. INTRODUCTION

Computed tomography is widely used for the evaluation of pulmonary abnormalities. CT images contain information about nodule size, shape, density, margin, and surrounding tissue. However, interpretation of CT examinations can be time-consuming, particularly when many examinations must be reviewed.

Deep learning methods have been investigated for detecting, segmenting, and classifying pulmonary nodules. Convolutional neural networks learn spatial image features directly from training data. Transfer learning can use representations learned from large image datasets, while attention and Transformer mechanisms can model relationships among different image regions.

The performance of a lung-cancer model depends on more than its architecture. Dataset composition, class definitions, preprocessing, patient-level splitting, data augmentation, and test-set design can substantially influence the reported result. Therefore, results from different studies should not be ranked solely by accuracy.

This review summarizes representative deep learning methods for CT-based lung-cancer classification and related segmentation tasks. It discusses datasets, architectures, evaluation metrics, validation limitations, explainability, and future research requirements.

II. REVIEW METHODOLOGY

A. Search Strategy

The review considered publications related to deep learning, lung cancer classification, pulmonary nodule analysis, CT imaging, and medical-image segmentation. Relevant search terms included “lung cancer classification deep learning CT,” “lung nodule classification CNN,” “lung cancer Transformer,” and “lung nodule segmentation CT.”

B. Selection Criteria

Studies were considered relevant when they addressed at least one of the following:

- CT-based lung-cancer or pulmonary-nodule analysis.
- Deep learning or hybrid computational models.
- Lung-nodule classification, detection, or segmentation.
- Quantitative evaluation using classification or segmentation metrics.
- Sufficient description of the dataset and experimental procedure.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

The review does not report percentages describing the proportion of studies using a particular method because a complete, independently verified extraction table was not available.

C. Extracted Information

The extracted information included:

- Model architecture.
- Dataset and imaging modality.
- Classification or segmentation task.
- Preprocessing and augmentation.
- Reported evaluation metrics.
- External or cross-dataset validation.
- Explainability and computational reporting.

III. Datasets

A. IQ-OTH/NCCD

The IQ-OTH/NCCD dataset contains CT scan slices grouped into normal, benign, and malignant categories. The dataset description reports 1,190 CT images from 110 cases, comprising 40 malignant, 15 benign, and 55 normal cases. Because multiple slices may originate from the same patient, patient-level splitting is important when evaluating models on this dataset [4].

Lung-Eff-Net was evaluated on IQ-OTH/NCCD for three-class classification. The study reported 99.10% accuracy and a ROC value between 0.97 and 0.99 on the test set [1].

B. LIDC-IDRI

LIDC-IDRI is a public CT dataset containing lung-nodule annotations from radiologists. It is used for nodule detection, segmentation, and classification research. Reader disagreement and the definition of reference labels should be reported clearly when using this dataset.

Wavelet U-Net++ was evaluated for lung-nodule segmentation using LIDC-IDRI and reported a mean Dice coefficient of 0.936 and a mean IoU of 0.878 [2].

C. LUNA16

LUNA16 is used extensively in pulmonary-nodule detection research and is derived from the LIDC-IDRI resource. Studies using LUNA16 should clearly state whether the task is nodule detection, malignancy classification, or another prediction problem.

D. Dataset Limitations

Datasets differ in scanner characteristics, slice thickness, reconstruction settings, patient populations, annotation procedures, and label definitions. Consequently, performance values obtained from different datasets are not directly equivalent.

IV. Deep Learning Architectures

A. Convolutional Neural Networks

CNNs are commonly used because they can learn local image patterns such as texture, boundary, and shape. In lung CT analysis, CNN features may represent visual characteristics associated with pulmonary nodules.

Lung-EffNet is a transfer-learning model based on EfficientNet with a modified classification head. The study evaluated EfficientNet variants B0 through B4 on IQ-OTH/NCCD. The reported best model was based on EfficientNetB1, with 99.10% test accuracy and ROC values between 0.97 and 0.99 [1].

These results represent the study's reported test-set performance. They do not establish equivalent performance on other datasets or in clinical practice.

B. Attention Mechanisms

Attention mechanisms can assign greater weight to selected channels or spatial regions. They are used to emphasize features that may be relevant to classification.

Attention maps can provide visual information about model behavior, but they should not automatically be treated as faithful explanations. Explanation quality should be evaluated using appropriate quantitative tests, such as perturbation analysis, stability testing, localization agreement, or expert review.

C. Transformer Models

Transformer architectures divide images or image features into tokens and model relationships between them. Swin Transformer uses hierarchical representations and shifted windows to model local and broader context.

Transformer performance depends on training data, pretraining, augmentation, optimization, and task definition. Reported results should therefore be interpreted together with the dataset and validation procedure.

D. Hybrid CNN–Transformer Models

Hybrid models combine CNN-based local feature extraction with Transformer-based contextual modeling. This combination may be useful because lung nodules contain local morphological characteristics and contextual information from surrounding lung tissue.

However, architectural complexity may increase parameter count, memory use, and inference time. Future reports should include computational measures in addition to predictive metrics.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

E. Segmentation Networks

Segmentation models identify the pixels or voxels corresponding to lungs or nodules. Segmentation may support classification by restricting analysis to an anatomical region and may improve visual interpretation.

Wavelet U-Net++ combines U-Net++ with wavelet-based processing for lung-nodule segmentation. The reported mean Dice coefficient was 0.936 and the mean IoU was 0.878 on LIDC-IDRI [2]. These are segmentation results and should not be compared directly with classification accuracy.

V. Reported Study Results

TABLE I
Representative Reported Results

Study	Model	Dataset	Task	Reported Result
Raza et al. [1]	Lung-EffNet / EfficientNet	IQ-OTH/NCCD	3-class classification	99.10% accuracy; ROC 0.97–0.99
Agnes et al. [2]	Wavelet U-Net++	LIDC-IDRI	Nodule segmentation	Dice 0.936; IoU 0.878
Yuan et al. [3]	Meta-analysis	Multiple datasets	Lung-cancer diagnosis	Sensitivity 0.86; specificity 0.86; AUC 0.92

The values in Table I are not directly comparable. The first two rows are results from individual studies, whereas the third row reports pooled estimates from a systematic review and meta-analysis.

VI. Evaluation Metrics

Accuracy is not sufficient for evaluating an imbalanced medical dataset. Researchers should report sensitivity, specificity, precision, F1-score, receiver operating characteristic area under the curve, precision–recall area under the curve, confidence intervals, and calibration where appropriate.

For binary classification, sensitivity is defined as:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives.

For segmentation, Dice and IoU are defined as

$$Dice = \frac{2|P \cap G|}{|P| + |G|}$$

$$IoU = \frac{|P \cap G|}{|P \cup G|}$$

where P is the predicted segmentation and G is the reference segmentation [5].

Segmentation studies may also report boundary measures such as Hausdorff distance and average symmetric surface distance.

VII. Validation Challenges

A. Patient-Level Data Splitting

When multiple slices originate from the same patient, images from one patient must not be distributed across training and test sets. Otherwise, the test set may not represent independent patient-level evaluation.

B. External Validation

Internal test-set performance does not establish generalization to another hospital, scanner, imaging protocol, or patient population. External validation should be performed using data that were not involved in model development.

C. Dataset Shift

Changes in scanner manufacturer, slice thickness, reconstruction kernel, intensity range, demographic distribution, and disease prevalence can affect model performance. These factors should be documented and analyzed.

D. Calibration and Uncertainty

A model may produce high classification accuracy while generating unreliable probabilities. Calibration analysis and uncertainty estimation can help identify predictions that require additional review.

E. Explainability

Saliency maps, Grad-CAM, attention maps, and related methods may help visualize model behavior. However, the presence of a heatmap does not prove that the explanation is accurate or clinically meaningful.

F. Computational Reporting

Studies should report parameter count, FLOPs, memory requirements, inference time, and the hardware used for evaluation. These details are necessary for reproducibility and deployment assessment.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

VIII. Discussion

The reviewed literature includes CNN-based, transfer-learning, attention-based, Transformer, and segmentation approaches for CT-based lung-cancer analysis. Lung-EffNet reported high internal test performance on IQ-OTH/NCCD, while Wavelet U-Net++ reported segmentation results on LIDC-IDRI [1], [2].

The 2025 meta-analysis included 315 studies. Among these, 209 studies addressing lung-cancer diagnosis reported pooled sensitivity of 0.86, specificity of 0.86, and AUC of 0.92. These pooled values provide a broader evidence summary than the result of any single benchmark study [3].

The evidence does not justify claiming that one architecture is universally superior. Differences in dataset, task, label definition, preprocessing, and validation design prevent direct ranking across studies. High performance on a public dataset should therefore be presented as study-specific internal performance.

IX. Future Research

Future research should focus on:

- Patient-level data splitting and transparent dataset descriptions.
- External validation using independent institutions.
- Cross-dataset testing to measure domain shift.
- Standardized reporting of sensitivity, specificity, AUC, calibration, and confidence intervals.
- Quantitative evaluation of explanation faithfulness and stability.
- Uncertainty estimation for identifying difficult cases.
- Reporting of model size, computational cost, memory use, and inference time.
- Prospective evaluation before claims about clinical usefulness are made.

X. Conclusion

Deep learning methods have been applied to CT-based lung-cancer classification and lung-nodule segmentation. Representative studies have reported strong results on selected datasets, including the Lung-EffNet test-set results on IQ-OTH/NCCD and the Wavelet U-Net++ segmentation results on LIDC-IDRI [1], [2].

These results are study-specific and should not be interpreted as proof of external generalization or clinical readiness. A 315-study meta-analysis reported pooled diagnostic sensitivity of 0.86, specificity of 0.86, and AUC of 0.92. Future work should emphasize independent validation, patient-level evaluation, transparent reporting, uncertainty estimation, calibration, explainability, and prospective clinical assessment [3].

References

[1] R. Raza et al., "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," Engineering Applications of Artificial Intelligence, vol. 126, Art. no. 106902, 2023.

[2] S. A. Agnes, A. A. Solomon, and K. Karthick, "Wavelet U-Net++ for accurate lung nodule segmentation in CT scans: Improving early detection and diagnosis of lung cancer," Biomedical Signal Processing and Control, vol. 87, Part A, Art. no. 105509, 2024.

[3] X. Yuan et al., "Systematic review and meta-analysis of artificial intelligence for image-based lung cancer classification and prognostic evaluation," npj Precision Oncology, 2025.

[4] Iraq-Oncology Teaching Hospital/National Centre for Cancer Diseases, "The IQ-OTH/NCCD lung cancer dataset," Mendeley Data, dataset description, 2023.

[5] A. S. Al-Shamayleh, M. Qatawneh, and H. A. Elsalamony, "Advanced Retinal Lesion Segmentation via U-Net with Hybrid Focal–Dice Loss and Automated Ground Truth Generation," *Algorithms*, vol. 18, no. 12, p. 790, Dec. 2025.